

**DATA MINING SYSTEM, METHOD AND APPARATUS FOR INDUSTRIAL  
APPLICATIONS**

5 [0001] This application claims the benefit, for purposes of priority under 35 U.S.C. §  
119(e), of U.S. Provisional Patent Application Number 60/266,640, filed February 6,  
2001.

**FIELD OF THE INVENTION**

10 [0002] The invention relates to data mining and, more particularly, to a system  
method and apparatus for mining data and providing services related thereto using a  
communication system, such as the Internet, in industrial applications.

**BACKGROUND**

15 [0003] Data mining explores detailed business transactions to uncover patterns and  
relationships contained within a particular business activity and history. Data mining  
can be done manually by "slicing and dicing" the data until a pattern becomes  
obvious. A term of art in the field, "slicing and dicing" refers to the ability to move  
between different dimensions of warehoused data. It can be performed with  
programs that analyze the data automatically. Using computer technology to look  
for hidden patterns in a collection of data, data mining for marketing research, for  
example, might reveal that customers interested in one product will also be  
20 interested in another. In other areas, data mining can be useful in scientific  
research, economics, criminology, and many other fields. In general, there exists  
specialized database software for data mining.

25 [0004] A data warehouse is a database designed to support decision making in an  
organization. It is batch updated and can contain enormous amounts of data.  
Hence, the moniker "warehouse". For example, large retail organizations can have

100GB or more of transaction history. When the database is organized for one department or function, it is often called a "data mart" rather than a data warehouse.

[0005] The data in a data warehouse is typically historical and static and may also contain numerous summaries. It is structured to support a variety of analyses,

5 including elaborate queries on large amounts of data that can require extensive searching. When databases are set up for queries on daily transactions, they are often known as operational data stores (ODSs) rather than data warehouses.

[0006] At present, there exist no data mining services relating to industrial applications that provide data mining and data mart services as described above.

10 Until now, data services for industrial applications have been offered by consultants who manually compile and audit data on a case-by-case basis. Such manual compilation may be inadequate to provide a complete or accurate analysis of the industry being researched. Manual accounting methods, moreover, cannot provide a continuous or virtually continuous update of the industry and are subject to human  
15 error.

### Summary

[0007] The long felt, but unmet, needs described above are addressed by various aspects of the system and method according to the present invention.

One aspect of the present invention, for example, provides a method for retaining a  
20 client subscription to an on-line site that provides access to industrial products and services, including data mining services in connection with a data warehouse populated with data relating to at least one industrial application. According to the method, the on-line site provides access at least to industrial products and services. A subscription is offered to the client to access the on-line site. If the client accepts  
25 the subscription offer, the client is further offered at least some access to the data mining service free of charge, though is otherwise charged a fee for subscribing to

10068614-020602

the on-line site. The client subscription is thereby retained by rewarding the client with the data mining analyses service access free of charge.

10056614-626602

5 [0008] In another aspect of the present invention, a method in a data mining system provides analysis services directed to industrial control-related data originating at a plurality of client industrial systems. The data mining system is in communication with the plurality of client industrial systems, is accessible over a network by a user, is associated with a data warehouse, and comprises at least one data mining application. The method comprises several steps. Industrial control-related data is  
10 collected from the plurality of client industrial systems. The collected industrial control-related data is stored in the data warehouse associated with the data mining system. User access over the network to the at least one data mining application is provided. In response to the user-directed data mining application, data is retrieved from the data warehouse and processed. Processed data is then delivered over the  
15 network to the user.

[0009] In yet another aspect of the present invention, a method is provided for delivering industrial control-related on-line services to a user during an on-line session. The method may comprise the following steps. The user is provided with  
20 access over a network to a data mining system during the on-line session. The data mining system comprises at least one application in communicatoin with a data warehouse. The data warehouse comprises data collected from among network-delivered data originating with a plurality of industrial control systems. The application allows the user to conduct analyses of data in the data warehouse and to  
25 view the results of the analyses. The user is also provided, during the same on-line session, with access to non-data mining industrial control-related content, the access provided during the same on-line session.

[0010] In a further aspect of the present invention, a method is provided for  
30 generating a data structure comprising industrial control-related content for

presentation to a user over a network during an on-line session. A document is generated for presentation over the network to the customer during the on-line session. A first link is inserted into the document that, when selected by the user, points to a second document relating to an industrial-control related data mining service. A second link is inserted into the document, such that, when selected by the user, it points to non-data mining industrial control-related on-line content. The user is thereby provided access to the industrial control-related data mining service as an incentive to also access the non-data mining industrial control-related on-line content.

[0011] In still another aspect of the present invention, a data mining system provides analysis services directed to industrial control-related data originating at a plurality of client industrial systems. The data mining system is in communication with the plurality of client industrial systems and is further accessible over a network by a user. The data mining system is associated with a data warehouse and comprises at least one data mining application. The system comprises data collection means adapted for collecting data from the plurality of client industrial systems, a data warehouse coupled to the data collection means and adapted for storing data collected from the plurality of industrial systems, on-line analytical processing means coupled to the data warehouse and adapted for analyzing industrial systems data, and user-interface means for presenting to the user the results of on-line analytical processing.

[0012] Various other aspects of the system and method according to the present invention are illustrated, without limitation, in the figures, the description below, and the appended claims.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Figure 1 is a diagram of an embodiment of a system according to the present invention.

5 [0014] Figure 2 is a flow diagram of an embodiment of a method according to the present invention.

[0015] Figure 3 is a diagram of an embodiment of a system according to the present invention, illustrating real-time analysis.

[0016] Figure 4 is a diagram of an embodiment of a business model according to the present invention.

## 10 DETAILED DESCRIPTION

[0017] Figure 1 shows a system diagram 100. One or more remote sites 102, in this instance industrial plants or sites 102a-102c, are coupled to an on-line communication network 104, such as the Ethernet, Internet, Intranet, local area network or the equivalent. The connection between the remote sites 102 and the  
15 communication network 104 is accomplished in accordance with any of the known protocols including, but not limited to, TCP/IP, ModBus, etc. Information or data is collected by a collection mechanism 106. The collection mechanism 106 in one aspect is performed automatically by, for example, software. This may be implemented using any of the known scripting languages, such as HTML or Java.

20 [0018] The data collected by the collecting mechanism 106 is stored in data warehouse 108. On-Line Analytical Processing (OLAP) 110, sometimes referred to as "multi-dimensional analysis," analyzes the data stored in the data warehouse 108 according to a predetermined analysis routine. Such routine may be implemented automatically by software. A report 112 is generated according to predetermined  
25 processing. In one aspect, the processing, which may be software driven, provides a user-specific report 112a in a predetermined format, such as a best practices chart

112b. As illustrated, the best practices chart may plot the efficiency of a particular piece of equipment, such as an industrial machine. As shown in Figure 1, the best practices chart plots efficiency versus the firing rate of the equipment.

5 [0019] In one implementation, the remote sites 102 are the clients, i.e., users. A client may be an industrial manufacturer that may wish to measure the efficiency of its plant or its equipment, for example, a boiler. Data collected from boilers at remote sites 102a-c are uploaded through the communication network 104. The information is collected by collection mechanism 106 and stored in data warehouse 108. The data stored in the data warehouse is data to be mined by the OLAP 110.  
10 In one aspect, the OLAP 110 is designed according to information entered on the specific type of equipment and model, boilers in this example, regarding the efficiency of such equipment.

15 [0020] The data is sliced into data marts (or markets) according to predetermined parameters set either by the user or the OLAP 110 software programmers, i.e., the service provider. For example, the data may be sliced into machines of a particular type, such as boilers. Such a cross-section, from equipment having a common characteristic, provides the user with superior data points with which to compare the user's own equipment. An example of such a common characteristic might be the external temperature of the equipment. This would populate the data market with  
20 information on equipment in similar temperature environments to determine, for example, the efficiency of equipment when this external factor puts a load on the equipment. External temperature may be obtained using isothermal maps available off the Internet. Other common factors, such as the age of equipment, geographic location and industry usage may alone, or in combination, be factors taken into  
25 account in accordance with the present invention.

[0021] The data market may comprise data with any predetermined characteristic including data collected from equipment having dissimilar features or used in

disparate contexts. For example, it such a boiler in a warehouse, a school or a food processing plant. The user may use the data market to ascertain a cross-section of information regarding the efficiency of boilers more effectively than with the existing method of manually compiling information provided by consultants.

5 [0022] With reference to the flow diagram shown in Figure 2, the remote sites 202  
(102, in Figure 1), may be one or more locations (e.g., industrial operations)  
(202a...202n). The locations are in communication with a central database, via a  
communication network 204. Scanner 206 periodically or continuously connects to  
one or more of the remote sites to collect data. The scanner may do this either  
10 automatically or at the prompting of the service provider. In addition, the user may  
upload information at their discretion. In the boiler example, the scanner may  
download fuel consumption data or steam output information, for example. Of  
course, the scanner may download any pre-selected information that is selected for  
populating the data warehouse.

15 [0023] A Data Transformation Service (DTS) 208 mechanism scrubs the data that  
has been scanned and uploaded by the scanner 206. Data scrubbing is a well  
known technique that filters incoming data so that unnecessary data is removed and  
does not waste data warehouse space. The DTS, for example, may remove the  
internet routing address or other data that may not be useful for determining the  
20 efficiency of the boiler in the above example. The scrubbed data is then stored in  
data warehouse (D/W) 210. The Online Analytical Processing (OLAP) mechanism  
212 analyzes and processes the data in the warehouse 210. The data may be  
sorted and redistributed into smaller data warehouses called data markets, or marts.  
These data marts may include data representative of specific characteristics of the  
25 piece of equipment under consideration.

[0024] The DTS 208, Data Warehouse mechanism 210 and OLAP 212 may be  
implemented using off-the-shelf software, hardware, firmware or a combination of

them. Software and hardware for the DTS 208, Data Warehouse and OLAP 212 may be procured through well-known public vendors or original equipment manufacturers (OEM). For example, the data warehouse hardware and software is publicly available from Oracle Sequel 2000. The OLAP processes may presently be  
5 obtained from Knosis, Inc. of Boise, Idaho, e.g., its Knosis product, which may be launched using Microsoft Excel 2000. Other vendors include Applix, Inc. of Westboro, Massachusetts, Brio Technology, Inc. of Palo Alto, California, and Business Objects of San Jose, California.

[0025] Next, a chart 214, in this case a best practices chart, is created based on the  
10 result of the on-line analytical processing 212. In Figure 2, the best practices chart may plot efficiency versus firing rate for boilers from which data has been collected for two customers. Data collected for each of the two customers is plotted relative to an ideal best practice curve. As shown in the figure, the plot for customer 1 exceeds the ideal best practice at lower firing rates, but falls below it at higher rates. This plot  
15 for customer 2 lies below the ideal best practice curve for all firing rates.

[0026] The industrial process best practice analysis of the present invention may be instantaneous as well as continuous, since the data from the various remote sites may be scanned and updated to the data market for analysis virtually immediately. Any combination of factors can be selected to determine the choice of data, the  
20 characteristics of the data to be presented and the manner in which the data is presented to the client, i.e., user. The present invention provides confidentiality to the users because the users access the best practices information through the Internet. It is a simple matter to maintain confidentiality of the user located at the far end.

[0027] The best practices chart and the OLAP 212 processing will now be described  
25 in more detail with respect to the diagram 300 shown in Figure 3. A boiler 302, for example, is equipped with sensors to detect operating parameters such as the total

20060614 090602  
20060614 090602



fuel used or the total steam produced. That information is uploaded, via the communication network, to the data warehouse. The information is "sliced" or otherwise processed, by the scanner 206, DTS 208, or other means and is stored in the data warehouse 304. In this instance, the data warehouse 304 is logically

5 organized according to the arrangement shown in Figure 3, wherein the data is collated according to equipment number 304a, i.e., boiler number, and one or more parameters relating to the respective equipment, i.e., total fuel 304b or total steam 304c. The known efficiency equation that relates steam to fuel is used to generate the efficiency number 304d that is stored, as well, in the data warehouse 304. Other

10 equations or relationships could as easily be used. In addition, the data warehouse 304 may be organized as multi-dimensional abstract space, the dimensions of which are defined by two or more variables. In the example shown in Figure 3, the data warehouse may be arranged according to a 3-dimensional space 304e. In another example, a multi-dimensional space may comprise dimensions corresponding to

15 time, location and equipment parameters. Arranged in this manner, the OLAP 212 sorts and processes the data according to one or more of the dimensions.

[0028] As shown in the figure, OLAP 212 may process the data in the warehouse (or data mart) 304, thereby providing a real time analysis. In the figure, multiple types of

20 OLAP processing charts 306a to 306c are shown. Each of the charts represents a different efficiency response curve for respective pieces of equipment. For example, chart 306a illustrates a typical response curve, wherein the equipment functions with a predicted and continuous efficiency over a range. Chart 306b illustrates an atypical problem scenario, wherein the plotted equipment efficiency data briefly

25 spikes at a particular point as a result of an equipment problem. Chart 306c illustrates a situation in which the equipment operates nearly constantly over a period of time. OLAP 212 can arrange and manipulate the data to allow the user to flexibly view the data from particular points of view. For example, the OLAP 212 can arrange the data according to geographic region, allowing for prediction of how

30 processes might be run in relation to a particular feature common to the region, such

20060614.020600

5

10

15

25

or in addition, use of one or more of the applications, products or services may incur a charge. The applications, products or services may be provided on a surcharge or commission basis. For example, use of services provided by a client may incur a commission fee in return for providing the forum in which the clients meet and  
5 arrange to agree upon the thing to be exchanged.

[0032] In another aspect of the business plan, the clients are provided free data mining services. This may include free access to the OLAP resource 212. That is, the client will be able to access the data warehouse information and employ the OLAP resource 212 to slice the data in any manner that the client desires. This is  
10 done, for example, to obtain best practice information for industrial equipment, such as boilers. In return for such services, the client may pay a subscription fee for the on-line forum. In the alternative, the web site may provide such free services in order to entice users to view and/or visit advertisements and/or web sites relating to the content of the web site, i.e., industrial applications.

[0033] In other words, the business model may be characterized as customer driven: the customer provides the information and, indeed, may provide it to other customers. This differs from the traditional method of providing manual consulting services, in which a vendor provides and mines the data for the client. The business model of the present invention may include other features as well. A proactive  
15 business model, i.e., customer or service driven, may provide periodic reports to the user. These reports may be in the form of weekly or monthly reports. The periodic reports may have a predetermined format designated by the user. The report may be formatted to best communicate the information to the client according to the user's particular industrial operation.  
20

[0034] Another aspect of the business model is to provide a performance guarantee service. In other words, in order to incentivize the purchasing of products and services provided over the web site, the business model offers a guarantee on the  
25

performance of such product or service. According to a further implementation of the business model, the incentivization program may be provided as part of the subscription in order to attract users to the web site and entice them to become long term members. In the alternative, users may purchase the guarantee on a per-item basis.

[0035] Another feature of the business plan is to provide inventory control for the user. In more detail, the user is provided with software applications on-line that provide optimized maintenance, deliveries, etc. For example, the inventory control allows the user to ensure that the optimal amount of equipment is supplied to the user remote site at any instant in time. This encourages long-term use of the website, because a will come to depend on the optimal inventory control.

[0036] Another aspect of the business plan is to provide preventative maintenance control in order to incentivize long-term subscription to the web site. The web site provides reports and predictions of equipment failure based on the data analysis performed by the OLAP 110 on the data in the data mart. Further, the model may include a crisis response, such as alarm and/or dial out capability for automatically alerting the user to an actual or potential crisis identified automatically by the OLAP 110, sensors, etc. In addition, there may be provided the service of reporting historical operations of the user's equipment or operation. Such a service is useful, from a business point of view because it addresses user's business need to perform such reporting, as is normally performed by quality control departments.

[0037] Each of the above-described features of the business model are sources of revenue and can create and incentive for a user to retain the subscription to the web site for a relatively long period of time. Further, it is by design that the business plan encourages the user to return to the site frequently to monitor the user's operations and to lock the user in so that the user need not seek any other source for product or service, particularly relating to the industrial field. In essence, the business plan of

the present invention provides a one-stop shopping forum that satisfies all or nearly all of the user's industrial application and product needs.

- [0038] In addition to the embodiments of the aspects of the present invention
- 5 described above, those of skill in the art will be able to arrive at a variety of other arrangements and steps which, if not explicitly described in this document, nevertheless embody the principles of the invention and fall within the scope of the appended claims.

20000414-020602